

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is a publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/28061>

Please be advised that this information was generated on 2017-12-05 and may be subject to change.

# IDIOMATICITY AND TERMINOLOGY: A MULTI-DIMENSIONAL DESCRIPTIVE MODEL

*Henk Barkema*

*Abstract.* In this article we discuss the traditional terminology for idioms and other lexicalised expressions. We claim that it is inadequate for accurate, explicit and detailed descriptions of the various linguistic characteristics of expressions. As an alternative, we introduce a well-defined model with which three different types of linguistic characteristics can be captured with a small set of labels: compositionality, collocability and flexibility. What is new in our approach is that a *set* of labels is given to one expression instead of only one label. Each of these either indicates how the expression scores on a given scale or indicates additional characteristics of the expression (e.g. idiosyncratic syntax). In an appendix we provide an inventory of terms discussed in this article, with their definitions.

## 1. Introduction<sup>1</sup>

The introduction of new terminology for a topic that has been explored quite thoroughly already is a risky enterprise. By adding a new set of terms to one which has become firmly established, one inevitably runs the risk of causing confusion where one wanted to bring about clarity. However, the various traditional classifications of idioms and other lexicalised expressions in articles and handbooks of grammar are generally used to categorise these expressions under single labels. In this way the impression is given that expressions are homogeneous in their characteristics.<sup>2</sup> In our opinion, however, the traditional one-dimensional classifications can only be used to indicate which expressions have *more or less* the same characteristics, because generally the types of characteristics of lexicalised expressions are only loosely related. It is true that practically all such expressions are more or less limited in their morpho-syntactic freedom, that generally substitution of lexical items in these expressions is limited to some extent and that there are many with idiosyncratic semantic characteristics. But if one tried to create classes with *exactly* the same characteristics, the result would be a number of classes that is far greater than the traditional set of three, which distinguishes between idioms, restricted collocations and free constructions. To illustrate the weakness of the present classifications, we discuss two examples.<sup>3</sup>

<sup>1</sup> I would like to thank Wietske Vonk, Peter-Arno Coppen, Martin Everaert and the anonymous referee of *Studia Linguistica* for their valuable comments on earlier versions of this paper. I am very grateful to Jan Aarts; many parts of this article are the results of discussions we had together.

<sup>2</sup> See also e.g. Fernando & Flavell (1981: 17 ff.).

<sup>3</sup> The definitions of words and expressions in this article are from *Longman Dictionary of English Language and Culture* (Summers 1992, henceforth *LDOELC*), unless other references are given.

- (1) *the bottom/first rung of the ladder* ('the lowest level in the organization')

Expressions like this one are traditionally classified as 'idioms' because of their non-compositionality. Alternative terms are 'figurative idiom' (Cowie 1983) and 'non-compositional metaphor' (Wood 1986). Cowie uses the term 'figurative idiom', because people will generally be able to imagine a plausible relation between the idiomatic and the literal meaning of the expression. Although the expression is non-compositional, Wood (1986) does not use the term 'idiom' because lexical substitution is possible: *bottom* can be replaced with synonyms, antonyms, near-synonyms, etc. (see also section 2.1). Such substitutions are not possible in an idiomatic expression as *hook, line and sinker* ('with expressions of belief: completely'). The present terminology cannot adequately label these different degrees of lexical freedom. Finally, the expression can take various different syntactic forms, for example *ladder* can be preceded and followed by modifiers. This is not possible in *hook, line and sinker*, which can take one form only. This difference in morphosyntactic flexibility cannot be expressed with the traditional terminology either.

- (2) *bed and breakfast* ('(a private house or small hotel that provides) a place to sleep for the night and breakfast the next morning for a fixed price')

This expression is different from example 1 because the meanings of its lexical items form part of the meaning of the expression. It is traditionally classified as 'binomial' (see Quirk et al. 1985) or 'collocation' (see Wood 1986, and Cowie et al. 1983). The term 'binomial' highlights the presence of two nouns in the expression, while 'collocation' focuses on its limited lexical freedom. These terms do not show that the expression has an institutionalised meaning which is to some extent idiosyncratic, because it is more than the sum of the meanings of the lexical items in (and the syntactic structure of) the expression (see the definition of example 2). This is not the case with *Just Married* for example, which has a special pragmatic function, but no idiosyncratic meaning. No labels are available to express the difference in meaning between this expression and example 2; nor are there labels to indicate that *bed and breakfast* is more morphosyntactically frozen than, for instance, *developing country*, which can be pluralised. Finally, although *breakfast* in example 2 can be replaced by *board*, no further lexical variation is possible, as in *day after day*, in which *day* can be replaced with near-synonyms as *night*, *second*, *month*, *year*, etc., while it has more freedom than e.g. *wishful thinking*, which is lexically invariable. Again: no terms are available to indicate this difference.

On the basis of these two examples it is safe to conclude that the variety of characteristics among lexicalised expressions is so great that if one used



the traditional classifications, one would be bound to sweep a lot of valuable information under the carpet. It is true that the various features can be *described* in terms of degree of compositionality, and of collocational and morphosyntactic freedom, as we have just done. However, such descriptions lack conciseness and precision.

In this paper we introduce a multi-dimensional model which can be used for much more explicit and precise classifications of lexicalised expressions than have so far been possible with the traditional terminology. Another advantage of the model is *uniformity in description*, because the characteristics of an expression are pinpointed with a fixed set of parameters. Although traditional classifications can only be used for the description of types, our model can also be used for tokens. Not only can one describe whether a token is a base or variant form, but also, if it is a variant form, which types of variation it has.

## 2. Discussion of the literature

### 2.1. *Idiom*

For more than a century the majority of linguists have used the term 'idiom' to refer to lexicalised expressions with idiosyncratic meanings.<sup>4</sup> An example is Henry Sweet's (1889:139) definition:

The meaning of each idiom is an isolated fact which cannot be inferred from the meaning of the words of which the idiom is made up.

This can be regarded as the standard definition of the term and, generally speaking, later definitions boil down to the same two things: (a) idioms are expressions which contain at least two lexical items and (b) the meaning of an idiom is not the combinatorial result of the meanings of the lexical items in the expression. A more recent example of such a definition is:

... I shall regard an idiom as a constituent or series of constituents for which the semantic interpretation is not a compositional function of the formatives of which it is composed (Fraser 1976:103).

Although practically all definitions of idioms require them to be 'lexically complex', an exception to this rule is Charles Hockett's definition (1958:172), which implies—as Hockett admits—that words and morphemes can be idioms too and that all forms which are not idioms themselves, contain them:

Let us momentarily use the term 'Y' for any grammatical form the meaning of which is not deducible from its structure. Any Y, in any occurrence in which it is not a constituent of a larger Y, is an idiom.

In this sense idioms are the basic semantic units of a language, no matter whether they are morphemes, words, phrases, clauses or sentences. As the

<sup>4</sup> See also Wood (1986).

definition is not restricted to expressions with at least two lexical items, however, it has been rejected as impracticable by the majority of linguists.<sup>5</sup>

To be able to distinguish between expressions which contain both idiomatic and non-idiomatic parts (such as *rain cats and dogs*) and ones that are wholly idiomatic, additional terms have been introduced. For example, van der Linden (1992:8) distinguishes between 'idiomatic expressions' and 'idioms':

Idiomaticity is a property of aspects of the meaning of complex (multi-lexemic) expressions. Idiomaticity implies that these aspects are exclusively a part of the meaning of the expression as a whole.

An idiomatic expression is an expression some aspect(s) of the meaning of which is (are) subject to idiomaticity. Idioms are expressions all aspects of the meaning of which are subject to idiomaticity.

In our opinion, there is nothing wrong with well-defined classifications that are focused on one linguistic characteristic of lexicalised expressions (in this case compositionality). However, occasionally linguists have included other types of characteristics in their idiom definitions. For example, in the following quotation the phrase 'wholly non-productive in form' means that lexical substitution (by synonyms, near-synonyms or antonyms) is impossible:

An idiom is a complex expression which is wholly non-compositional in meaning and wholly non-productive in form (Wood 1986:2).

Although Wood's (1986) description of compositionality is thorough and at the same time allows for different degrees, we regret the inclusion of productivity in her definition of 'idiom' because the two different types of characteristics of lexicalised expressions—non-compositionality and non-productivity—are not mutually inclusive. As a consequence of her definition an idiomatic expression as *throw in the towel* ('to admit defeat') cannot be regarded as such, because one of its lexical items (*towel*) can be replaced by another (*sponge*); instead it is classified by her as a 'non-compositional metaphor'. In this way the set of idioms is restricted to a *subset* of expressions that are generally regarded as idioms, not because of their non-compositionality, but as a result of another linguistic characteristic. In this case we would opt for a classification with two subsets of idioms, namely one containing 'wholly nonproductive' idioms and the other containing idioms which are only productive to some extent.

Some linguists have included another type of characteristic, namely the relation speakers can perceive between the literal and the non-compositional meanings of idiomatic expressions, sometimes called 'warranty' or 'motivation'.<sup>6</sup> If such a link cannot be found, an expression is

<sup>5</sup> See for similar approaches e.g. Fernando & Flavell (1981:23), or Williams (unpublished manuscript) who uses the term 'semantic idiom'.

<sup>6</sup> See Carter (1987) and Fernando & Flavell (1981:32). Geeraerts (1992) and Ullmann (1973) have adopted this term from de Saussure (1916) – see Fernando & Flavell (1981:28).

'opaque' or 'fully unmotivated'. The reason why they have added this characteristic is that they regard both compositionality and motivation as one and the same semantic characteristic; an example is Fernando & Flavell (1981:29 ff), who label a fully unmotivated expression as 'pure idiom'. In our opinion motivation and compositionality are two different types of characteristics: compositionality concerns the relationship between basic senses of the lexical items in an expression and the meaning of the whole expression on the other and is, therefore, not restricted to idiomatic expressions, but also concerns lexicalised literal ones (see example 2: *bed and breakfast*), whereas motivation depends on the judgement of speakers about the relations between the literal and the institutionalised metaphorical meanings of idiomatic expressions (see §3.2). In relation to this distinction Cruse (1986:44) distinguishes between the terms 'opaque' (a form of motivation) and 'translucent' (a form of compositionality).

Although linguists have often taken the view that constructions are either fully compositional or fully non-compositional, there are a number of lexicalised expressions that can be regarded as compositional *to a certain extent*. With the exception of Cruse (1986:39–40), this fact has received little attention in the literature. In section 1 we briefly discussed *bed and breakfast*, whose meaning is more than the combination of the meanings of *bed* and *breakfast* (see also §3.2).

## 2.2. Collocation

Other terms that have been used for a long time are 'restricted collocation' (Aisenstadt 1979:71) and 'frozen collocation' (Bäcklund 1976). This type of collocation differs from another type of construction—'free combination' (Benson et al. 1986:xxiv), 'free construction' (Cowie 1981 and Altenberg & Eeg-Olofsson 1990), 'free word-combination' (Aisenstadt 1979:71), 'unrestricted collocation' (Carter 1987:63) or 'free combination' (Quirk et al 1985, §16.3)—in that the substitutability of its lexical items is limited:

Restricted collocability in English typically takes the form of word-combinations the constituents of which are restricted in their commutability. Such word-combinations we term 'restricted collocations' (RCs) (Aisenstadt 1979:71).

However, this characteristic is typical of idioms too. For example, Carter (1987:58) describes idioms as 'non-substitutable' or 'fixed' collocations. Generally the collocability of collocations is much greater than that of idioms, but idioms like *a drop in the bucket/ocean* and *the highest/top/lowest/bottom rung of the ladder* allow several alternatives, while the restricted collocations *foot the bill*, *curry favour* and *catch one's breath* allow no substitution at all (see Cowie (1981:228), or Cruse (1986:41), who



refer to such constructions with the term 'bound collocation'). On the other hand there are expressions which are limited in their range of collocability but are classified as free constructions.<sup>7</sup> It is in fact not their collocational restriction that sets restricted collocations apart from idioms, but their special meaning characteristics. Aisenstadt (1979:71) describes the difference between restricted collocations (RCs) and idioms as follows:

RCs are not idiomatic in meaning: they do not form one semantic unit; their meaning is made up as the sum of the meanings of their constituents (Aisenstadt 1979:71).

It is true that such expressions do not form one semantic unit. However, such restricted collocations have special meaning characteristics. For example, Nicolas (1995:234) claims that e.g. *sharp turn* and *take a decision* are semi-compositional because in each expression one of the lexical items has a 'specific, "non-free" sense'. In the same vein Cowie et al. (1983:x) note that:

one word (i.e. in the case of two-word expressions) has a figurative sense not found outside that limited context. The other element appears in a familiar, literal sense (cf. the verb and noun, respectively, in *jog one's/sb's memory*) and the adjective and noun in a *blind alley*). Some members of this category allow a degree of lexical variation (consider, for instance, *a cardinal error*, *sin*, *virtue*, *grace*), and in this respect 'restricted' collocations resemble 'open ones'.

Because of their semantic characteristics Cowie et al. (1983:x) decide to use the term 'semi-idiom' for such expressions:

... one word may have a common, literal meaning, while the other has a specialised sense which may be difficult to grasp. Examples of such 'semi-idioms' are *foot the bill* and *sink one's differences* (where the first word in both cases has a figurative meaning).

In our opinion the term 'semi-idiom' is more appropriate than 'restricted collocation' because it is really the *semantic* characteristics of such expressions that are essentially different from those of idioms, and not the collocability characteristics. In this respect such semi-idioms are similar to ones like *rain cats and dogs* that van der Linden (1992:8) classifies as 'idiomatic expressions' because '... some aspect(s) of the meaning of which is (are) subject to idiomaticity'. The only difference between *foot the bill* and *rain cats and dogs* is the number of (lexical) items in the idiomatic part.

In line with the traditional classification of constructions, Aisenstadt claims that:

All word-combinations in present-day English can be divided into idioms and non-idiomatic phrases. The latter constitute the vast majority of the word-

<sup>7</sup> See Cowie (1981:226) and Greenbaum (1970:9).

combinations functioning in speech. They can in turn be subdivided into free phrases and RCs. (Aisenstadt 1979:71).

Cowie (1981) makes another division, namely one between free constructions on the one hand and idiomatic expressions (idioms and semi-idioms) on the other; for a comparison of the two classifications, see Alexander (1987:109). In this article we have a similar classification. In addition we introduce a separate classification for the collocability characteristics of constructions parallel to one for compositionality.

### 2.3. *Transformational deficiency*

A type of characteristic of lexicalised expressions that was more or less ignored until the sixties, but which has received a lot of attention since then, is their limited morpho-syntactic freedom or 'limited flexibility'.<sup>8</sup> In addition to limited compositionality and restricted collocability, limited flexibility is the third linguistic type of characteristic by which lexicalised expressions differ from other types of construction.

In the same way as restricted collocability is not a characteristic of idioms only, there are many (more or less) compositional lexicalised expressions with limited flexibility:

Although an examination of their transformational behaviour offers insights into idioms as syntactic objects, it does not throw much light on idiomaticity *per se*. Not only are transformational constraints insufficient to distinguish the idiomatic from the non-idiomatic, they also fail to distinguish between semi-literal idioms with affective idiomatic adjectives (*my dear lady, that wretched woman, you poor man*, etc.), semi-literal idioms with non-affective idiomatic adjectives (*white lie, blue film, red cent*, etc.), completely non-literal idioms (*white elephant, blue stocking, turn a blind eye*, etc.) and peculiarities of usage (*criminal law, an utter fool, an outright lie, a real hero, drunken quarrels, miscellaneous articles*, etc.). All the adjectives in these items share the feature that they can be used only attributively. Predicative use, nominalisations and inflection for degree do not generally apply to these adjectives. On the basis of transformational deficiency alone, a feature, incidentally, which contributes to the status of these items as institutionalised set phrases, they would be classed as idioms of equal status (Fernando & Flavell 1981:43, 44).

We agree with their conclusion that

... idioms are at best only a sub-class of all transformationally deficient structures. All idioms are transformationally deficient, but not all transformationally deficient structures are idioms. Therefore any attempt to set up syntactic deviance as the primary norm to establish idiomaticity fails, since it cannot distinguish between idioms and these other transformationally deficient structures. (p. 44).

<sup>8</sup> See e.g. the transformational linguists Katz & Postal (1963), Fraser (1970) and Newmeyer (1974). See also Weinreich (1969), Chafe (1970), Gazdar et al. (1985), Carter (1987), Healey (1968), Fernando & Flavell (1981), Pawley & Syder (1983) and Cowie et al. (1983).



As idiomaticity and limited flexibility are not mutually inclusive, it is necessary to introduce a separate dimension for flexibility, in addition to ones for idiomaticity and collocability. We admit that at first sight a multi-dimensional classification seems less convenient than the traditional system which applies one label to one construction. However, we believe it is more realistic, and better expresses the complex interaction between the types of characteristics of such constructions.

#### 2.4. *Institutionalisation*

The fact that idioms and semi-idioms are institutionalised is usually taken for granted. A number of linguists, however, have given explicit attention to the notion of institutionalisation. For instance, one of Makkai's (1972) five criteria for the identification of idioms is the extent to which a construction has become current.<sup>9</sup> Two examples of descriptions of this type of characteristic of idioms and semi-idioms are:

By the institutionalisation of idiom we mean the regular association in a speech community of a given signification with a given syntactic unit (a compound, a phrase or a sentence), such that the resulting expression is interpreted non-literally. In other words, part of the phenomenon of idiomaticity is the institutionalisation of an asymmetry between sense and syntax in the case of compound, phrasal and sentential idioms. (Fernando & Flavell 1981:44).

What makes an expression a lexical item, what makes it part of the speech community's common dictionary, is, firstly, that the meaning of the expression is not (totally) predictable from its form, secondly, that it behaves as a minimal unit for certain syntactic purposes, and third, that it is a *social institution* (Pawley & Syder 1983:209; the italics are present in the original text).

Carter (1987:59) uses the term 'fixed expression' to refer to constructions '... with different degrees of possible fixity or "frozenness", both syntactic and semantic'. He adds that:

... there is no real advantage in drawing strict lines between idioms and non-idioms or in treating collocations separately from idioms; instead it makes more sense to try to illustrate the different degrees of variability of fixed expressions (p. 135).

On the other hand there are many 'familiar utterances' (see Moore & Carling 1982) or 'speech formulae' (see Pawley & Syder 1983). Some examples from Pawley & Syder (1983:206) are: *Can I come in?*, *Are you ready?*, *I enjoyed that*, *Do what you're told*. Although such constructions are institutionalised, they cannot be regarded as ones with an 'asymmetry between sense and syntax' or as ones that are 'semantically and syntactically fixed'. It seems, therefore, that in this respect there are not only two types of construction (namely institutionalised constructions

<sup>9</sup> See also Fernando & Flavell (1981: 19).

which are, and ones which are not, fixed) but also two parallel types of process: a *socio-linguistic* one during which constructions gradually become more current, and a *linguistic* one during which they gradually become more semantically, syntactically and collocationally fixed.

### 3. Introduction of a multi-dimensional model

In our discussion of the literature we distinguished between four types of characteristics: institutionalisation, collocational restriction, idiomaticity and transformational deficiency. At the same time we noted that the relations between these characteristics are complex. As little is known about the *exact* interaction between the various types of characteristics, it is impossible to create classes on the basis of this interaction. For this reason we introduce a descriptive model in this section, which is fundamentally different from the traditional one: instead of providing each construction with one label which highlights only one striking feature, we classify expressions at different *dimensions of description*, thereby sketching a rich and transparent profile with no more than a few terms. In section 4 we discuss a number of examples, and provide them with multi-dimensional classifications.

In this paper we use the term 'construction' to refer to a phrase, clause or sentence with at least two grammatical function slots realised by lexical items other than the definite or indefinite article. Because constructions have at least two function slots, compounds such as the following are excluded from the class of lexicalised expressions: *ice-cream*, *baby carriage*, *sidewalk*, *shoulderblade*, *kneecap* and *sidecar*. These compounds have one function slot only, namely that of 'head of the noun phrase', as opposed to expressions like *blind alley*, which has two: 'pre-modifier' (*blind*) and 'head of the noun phrase' (*alley*).

Various linguists have stressed the scalar nature of linguistic characteristics. Examples are Fernando & Flavell (1981), Alexander (1987) and Carter (1987); for a detailed discussion: see Wood (1986). In our model each type of characteristic has its own scale. Because of their scalar nature, an infinite number of classes could be distinguished by each time choosing slightly different positions on each of the three scales. Each class would then be only slightly different from another. Obviously such a classification would be impracticable. Therefore we have divided each scale into three parts—two extremes and one middle section. For each scale we introduce terms for the different parts. We try to make these terms correspond as much as possible with the traditional terminology.

In our view the choice for a multi-dimensional model has two theoretical advantages: each classification is related to a specific dimension of description—syntactic, semantic or other, and is therefore unambiguous. At the same time complex relations between various types of characteristics can

be made explicit. A practical advantage is that with a few labels an accurate and detailed profile can be given of each construction.

In the field of lexicography a number of linguists have broken new descriptive ground in this respect; for example, in their dictionary Cowie et al. (1983:xxi–xxxvii) present the collocability characteristics of expressions and use separate grammatical codes for their morphological and syntactic characteristics. Also, the Collins COBUILD English Language Dictionary (Sinclair 1987) provides separate descriptions of the flexibility characteristics of lexicalised expressions.<sup>10</sup>

A linguist who has already distinguished between different scalar types of linguistic characteristics is Carter (1987:63–4). He briefly discusses three different scales: collocational restriction, syntactic structure and semantic opacity in the form of ‘... sets of continua with fixed points but several intermediate categories’; for examples of classes of lexicalised expressions and for references to articles in which these classes are discussed: see Carter (1987:59). In Carter’s classifications on the various scales there appear to be some snags. On his ‘collocational restriction’ scale he distinguishes between ‘unrestricted’, ‘semi-restricted’ and ‘restricted’ collocation, but he has an additional type—‘familiar collocation’—between restricted and semi-restricted. It is not clear to us why this type is necessary. Although it seems that his collocability scale has two extremes (namely ‘unrestricted’ on the one hand and ‘restricted’ on the other) he remarks that restricted collocations are ‘generally more fixed and closed’, thereby leaving unclear whether such constructions can only occur at the extreme, or also somewhere near it.

On his ‘syntactic structure’ scale we find both ‘irregular’ and ‘flexible’. It is true that constructions with idiosyncratic syntax are generally inflexible, but the reverse is not. For this reason we prefer a distinction between descriptions of the morpho-syntactic structures of constructions on the one hand, and descriptions of their (scalar) flexibility characteristics on the other. Carter’s third scale concerns ‘semantic opacity’: this scale seems to involve both compositionality and motivation characteristics—two types we prefer to keep separate (see section 2.1). Nevertheless the ideas behind these scales form a valuable foundation for the construction of a solid model. In the rest of this article we give a detailed description of various types of characteristics of constructions. In addition we introduce and define terminology which can be used by grammarians and lexicographers for the specification of the various (types of) characteristics (see the appendix for the definitions).

<sup>10</sup> See Sinclair (1987:1076): the entry of ‘PHR’.



### 3.1. Familiarity and lexicalisation

In English and other natural languages, constructions are generally created by combining two or more lexical items in a grammatical structure. We call them 'standard constructions'. Examples are:

- (3) *a friendly woman* ('a fully grown human female who is acting or ready to act as a friend')
- (4) *buy a house* ('to obtain a building for people to live in by giving money (or something else of value)')

On the other hand many constructions have become institutionalised. The process of institutionalisation is a *socio-linguistic* one: a (standard) literal or metaphorical construction gradually becomes more current within a language community because of its specific meaning and/or pragmatic function.<sup>11</sup> When it has a higher than average currency in the language community, we call it a 'familiar construction' or 'expression'. Examples are *paddle one's own canoe* or *bed and breakfast*, but also ones as *What time is it?* and *Do what you're told!* Pawley & Syder (1983:205) call such constructions 'memorized sequences':

These are strings which the speaker or hearer is capable of consciously assembling or analysing, but which on most occasions of use are recalled as wholes or as automatically chained strings.

From the start of the process of institutionalisation an expression can go through a gradual process of lexicalisation, which is *linguistic* in nature: the application of general grammar rules to the expression becomes more and more subject to restrictions, with as a result limited flexibility as well as limited collocability and/or compositionality. An example of an expression which is gradually becoming lexicalised is *a new world order*. Initially it was used by President Bush to refer to the global balance of power after the Gulf War in 1991, but one year later it was also used to refer to a new balance of power in more general terms, as is illustrated with the following quote: 'A new world order is emerging in Paris's gilded salons (*The Independent*, 30/7/1992).

These restrictions on generally applicable grammar rules are generally 'idiosyncratic', i.e. they cannot be formulated in terms of (general) restriction rules and it is impossible to pinpoint exactly under what conditions the grammar rules are restricted.

At the beginning of the process of lexicalisation the expression is still a 'free expression': the general grammar rules are not subject to idiosyncratic restriction when they are applied to it. As soon as the application of one type of (syntactic, semantic or lexical) rule to an expression becomes subject to idiosyncratic restriction, it can be regarded as a

<sup>11</sup> For a discussion of the term 'currency': see Aarts (1991:57–8).

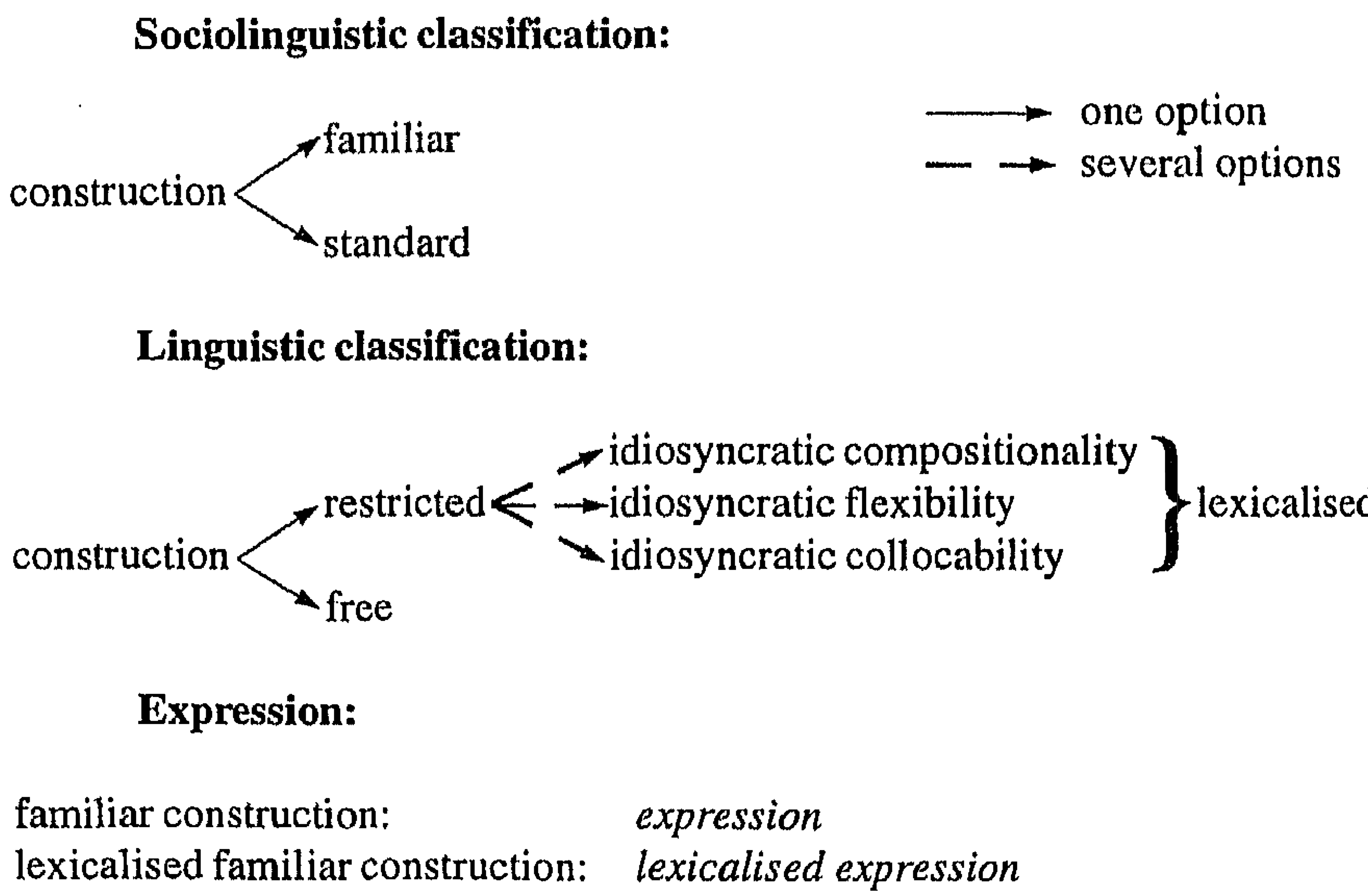
‘restricted expression’. Finally, when it has limited morpho-syntactic freedom and at the same time is collocationally restricted and/or no longer compositional, it has become a ‘lexicalised expression’.<sup>12</sup> Examples are:

- (5) *cup and saucer* (‘a cup and a saucer as a pair’)
- (6) *As thick as two short planks* (‘very stupid indeed’)

It should be noted that expressions do not necessarily go through a process of lexicalisation. For instance, *What time is it?* and *Do what you’re told!* are unlikely to become lexicalised. At the same time a large number of constructions are restricted only collocationally, and for this reason cannot be regarded as lexicalised. An example is:

- (7) *to work/operate/run a machine* (not e.g.: \**to manage a machine*, \**to conduct a machine*)

Combinatory dictionaries such as BBI (Benson et al. 1986) pay special attention to such constructions, while lexicalised expressions are dealt with in handbooks of grammar and dictionaries. The different types of construction can be classified in the following way:



3.2. *Compositionality*

The notion of compositionality is based on that of ‘sense’. We distinguish between three types: ‘basic’, ‘extended’ and ‘derived’. We use the term ‘sense’ for one or more concepts which can be denoted by one lexical item.

<sup>12</sup> In Barkema (1993, 1994a and 1994b) we used the term ‘received’ instead of ‘lexicalised’ because we did not yet distinguish between institutionalisation and lexicalisation.

The *basic sense* of a lexical item is the one from which other senses can be systematically derived, i.e. by means of semantic rules (see Aarts & Calbert 1979). It is its most obvious sense, that is: it is the first that comes to a speaker's mind when s/he comes across a lexical item as a single word, and not a text, sentence, clause or phrase containing it – for example when s/he sees it on a blank sheet of paper, or, to quote Burgess (1969:298): '... as words which carry meaning if chalked up singly or written in the sky by sky-writing aircraft'. It is often the most concrete sense, but not necessarily the one that is used most frequently. For instance, the basic sense of *house* is 'a building for people to live in'.

An *extended sense* of a lexical item is a temporary extension of its basic sense. It is the unique result of a violation of restriction rules. Such violation can only occur in context. As a result of the violation, sense adaptation rules are activated. There are three factors which play an important role in the activation of adaptation rules: (a) the basic sense of the lexical item; (b) the speaker's stereotyped beliefs about the referents of this basic sense, and (c) the syntactic environment in which it occurs.

This notion of violation is the central point in this description. In every syntactico-semantic construction lexical items impose restrictions on the selection of others. These selection restrictions are determined by the basic senses of lexical items. For example, such restrictions are violated in

- (8) But most Czechs believed until now that Slovakia would still be in favour of a federal *house*. (*The Economist*, 1/8/1992)

Here *house* no longer means 'building', but: 'country'. The result of the violation of restriction rules is an extension of the basic sense of *house*.

A *derived sense* of a lexical item is an extended sense which has become established. Like the basic sense, it can also be reproduced by a speaker when s/he comes across a lexical item as a single word. However, a derived sense is less obvious than a basic one; it generally is the second or following sense that comes to a speaker's mind when s/he is confronted with a lexical item as a single word. As opposed to extended senses, derived ones can be found in dictionaries. An example of a lexical item with a derived sense is *house*; one of its derived senses in LDOELC is: '(the members of) a law-making body, esp. when one of two'; an example is:

- (9) Sir Marcus divided *the house* and did not divide without organisation. Hughes's motion was roundly defeated. (*New Statesman & Society*, 2/8/1991).

From a diachronic point of view on language, derivation is a continuous process: gradually new senses are introduced into the language, while existing basic and derived senses may disappear, and derived senses can gradually become basic senses.

On the basis of the notions of basic, extended and derived sense we can



now describe 'compositionality'. The compositionality of a construction is the extent to which its meaning is the combinatorial result of the basic or derived senses of the lexical items in the construction and the syntactic relations in the constituent that contains these lexical items.

The meaning of a *standard literal construction* originates from a combination of the basic or derived senses of the lexical items in that construction and the syntactic relations in the constituent which contains these lexical items. A standard literal construction does not contain lexical items with extended senses. As its meaning is the combinatorial result of the senses and the syntactic relations we call it 'fully compositional'. Examples are:

- (10) *That girl is young* (basic senses only)
- (11) *That girl is a nasty cat* (basic and derived senses)

In example (11) the derived sense of *cat* ('nasty woman') is selected.

The meaning of a *standard metaphorical construction* originates from a combination of at least one lexical item with an *extended* sense and one or more lexical items with *basic*, *derived* or *extended* senses, and the syntactic structure of the constituent which contains these lexical items – exactly the same goes for metonymous and synecdochical constructions.<sup>13</sup> As is the case with standard literal constructions, metaphorical ones are fully compositional because their meanings are the combinatorial results of the senses of their lexical items and the syntactic structure of the constituents which contain them. However, as three types of sense can be combined instead of two, the compositionality of metaphorical constructions is more complicated than that of literal ones. An example of a standard metaphorical construction is:

- (12) These are early days for Mr Major, and some argue that his decision to put his most obvious potential successor, Kenneth Clarke, into the Home Office, *that graveyard* of political careers, shows he is not wholly self-confident. (*The Economist*, 1/8/92).

Some literal lexicalised expressions are fully compositional too. Some examples are: *Ladies and Gentlemen*, *if I may say so*, *putting it bluntly* and *having said that*. As a consequence we only find descriptions of the way in which they are used (not of their meanings) in dictionaries such as LDOELC or handbooks of grammar such as Quirk et al. (1985) (henceforth: CGEL), e.g.:

- (13) *Ladies and Gentlemen* ('USAGE: when speaking to a gathering of people')
- (14) *if I may say so* (style disjunct; see CGEL §8.124)

<sup>13</sup> For reasons of conciseness we use the term 'metaphorical' in contexts where it can be substituted with 'metonymous' and 'synecdochical'.

The meanings of some literal lexicalised expressions encompass more than the combinatorial results of the senses in their lexical items and their syntactic structures. In other words: the basic senses of the lexical items in these expressions play a role in, but form only parts of, their meanings. Take for example:

- (15) *bed and breakfast* ('a system of accommodation in a hotel or guest house in which you pay for a room for the night and breakfast the following morning', COBUILD)

Both *bed* and *breakfast* have their basic senses in this expression. However, these contribute only partly to the meaning of the expression, and other meaning properties such as *system of accommodation*, *pay*, *room*, *hotel or guest house* and *the following morning* cannot be inferred from them.<sup>14</sup> The same goes for e.g.:

- (16)  $X^{\text{time reference } Y} \text{ after } X^{\text{time reference } Y}$  ('continuously': e.g. *day after day*)

We call such expressions 'pseudo-compositional'. Some literal lexicalised expressions have pseudo-compositional meanings because the referents of their lexical items form a unit, as in:

- (17) *cup and saucer*

- (18) *man and wife*

Example 18 is pseudo-compositional for another reason: in the expression *man* means *husband*.

There is an important distinction between *standard* metaphorical constructions and ones that have become lexicalised, such as the following:

- (19) *put the cat among the pigeons*

- (20) *a thorn in one's side*

Standard metaphorical constructions have meanings which arise from extended senses (possibly combined with basic and/or derived senses). Rules for sense adaptation are activated as a result of violation of restriction rules. Lexicalised expressions which were originally standard metaphorical constructions, on the other hand, have established meanings that are not inferred from the senses of their lexical items. As opposed to many lexicalised expressions which were originally standard *literal* constructions, they do not have lexical items with basic senses that form (part of) their meanings. Therefore they are 'fully non-compositional'.<sup>15</sup> We use the terms 'idiomatic' and 'idiom' to refer to lexicalised expressions which are non-compositional.

Many idiomatic expressions have equivalents in the form of 'counterfeit form'. Such a form has the same syntactic form and contains the same

<sup>14</sup> See Makkai (1972) for the term 'idiom of encoding'. See Mel'čuk (1995) for the term 'quasi-idiom'.

<sup>15</sup> Makkai (1972) uses the term 'idiom of decoding' for such expressions.

lexical expression, but, because of the way in which it is used, has a meaning that is the combinatorial result of the meanings of the lexical items in the construction. An example of a counterfeit form is:

(21) *the helicopter circled round the tip of the iceberg*

Some lexicalised expressions are *partly idiomatic* (i.e. *partly non-compositional*), while they contain at least one lexical item with a basic sense. In the following examples the idiomatic parts are underscored:

(22) *a blind alley* ('a little narrow street with no way out at the other end')

(23) *rain cats and dogs* ('to rain very heavily')

Examples 22 and 23 are not pseudo-compositional because the basic or derived senses of only some lexical items in the expressions contribute to their meanings.<sup>16</sup> In pseudo-compositional expressions all basic or derived senses play a role in their meanings. We call examples such as 19 and 20, which only contain lexical items without basic or derived senses, *entirely idiomatic*.

In many lexicalised expressions the senses are 'dormant' instead of absent; in other words: the (basic) senses of idioms can be re-activated in certain contexts and by means of certain syntactic variations. An example is:

(24) He and the equally shark-like Gardini *pulled what looked like badly burned chestnuts out of the New Zealand fire* in the final of the Louis Vuitton Cup for the challengers. (*The Independent*, Sport, 9/5/92)

In section 2.1 we already referred to this notion with the term 'motivation'. Most (if not all) speakers of English are able to re-establish the relations between the institutionalised meanings of the idiomatic expressions *to play second fiddle* ('to play a less important part'), *to keep the ball rolling* ('to continue something, such as a conversation or a plan') and *to clip someone's wings* ('to prevent someone from being as active or powerful as before') and the compositional meanings of their (literal) counterfeit forms. The only requirement is the creation (or the imagination) of a context which corresponds with the institutionalised meaning of the idiom. In Cowie et al.'s (1983:xiii) classification such idioms are therefore labelled as 'figurative idioms'. The same is more difficult with expressions like *eat humble pie* ('to have to admit that one was wrong or that one has failed'), *by and large* ('on the whole, in general') and *red herring* ('a fact or subject which is introduced to draw people's attention away from the main point'). For this reason Cowie et al. (1983:xii) label such expressions as 'pure idioms'.

Some lexicalised expressions are idiomatic because they contain one or more lexical items which have no (basic, derived or extended) senses.

<sup>16</sup> Example 22 is polysemous; its second sense is entirely idiomatic (see §4, examples 77a and 77b).



These 'absent sense' lexical items only occur as part of lexicalised expressions. Some examples are: *spick and span*, *in lieu of*, *the whole caboodle*, *on behalf of*, *like wildfire*, *in fine/good fettle*, *at loggerheads with*, *kith and kin* and *on tenterhooks*. For instance, in LDOELC no separate senses are provided for the underscored lexical items. As each of these expressions contain at least one lexical item without a (basic) sense we call them 'absent sense' expressions. For such expressions speakers require special (etymological) knowledge to be able to recover the original basic senses. An example of an expression which was originally compositional is *by and large* which, according to Simpson & Weiner (1989) means 'to the wind (within six points) and off it'. The ability of speakers to re-establish relations between idiosyncratic meanings of idiomatic expressions and the meanings of their counterfeit forms on the one hand, and to recover the basic senses in 'absent sense' expressions on the other, depends on their knowledge, but also on their imagination! Therefore it is impossible to give a linguistic description of recoverability in terms of a relation between senses and meanings.

It is possible that there is a consensus among speakers about the recoverability of meanings of idiomatic expressions. In addition, it is possible that the recoverability of the meaning of an expression determines the degree of its flexibility. Whether or not this is true should be established through systematic examination of the knowledge and insights of a large number of speakers. As this leads us from the field of linguistics into that of psycholinguistics, however, we do not further elaborate on this topic. For an example of psycholinguistic research into the recoverability of the meanings of lexicalised expressions, see Van de Voort & Vonk (1995).

### 3.3. Flexibility

Lexicalised expressions minimally require the presence of specific lexical items and a specific syntactic structure for their meanings and/or pragmatic functions. For example, neither *your uncle* nor *and Bob* have a sufficient number of items to carry the idiosyncratic meaning of *and Bob's your uncle*. Therefore *and Bob's your uncle* is the *base form* of the expression.<sup>17</sup> The base form is the simplest morpho-syntactic form that the expression can take. Dictionaries such as COBUILD (Sinclair 1987) and LDOELC (Summers 1992) list base forms as main or sub-entries; examples are:

- (25) *rack and ruin* ('a ruined state, esp. of a building, caused by lack of care')
- (26) *and all that jazz* ('and other things like that')

<sup>17</sup> See Jespersen (1924), Healey (1968: 72) and Tovenia (1992).

Of the majority of base forms, all the grammatical function slots are filled, for example:

(27) *tip of the iceberg*

(28) *sitting duck*

Another set consists of base forms which have open function slots. Examples are:

(29) *what the X<sup>noun</sup>* ('used with various words, such as *hell*, *devil*, *blazes*, etc. when asking angry or surprised questions: what')

(30) *X<sup>time reference</sup> Y after X<sup>time reference</sup> Y* ('continuously'; e.g. *day after day*)

Everaert (1993) uses the term 'idiomatic pattern', Fillmore et al. (1988) the term 'formal idiom'. Pawley & Syder (1983:211) give examples of 'sentence stems' with open slots such as *NP be-TENSE sorry to keep-TENSE you waiting* and *Who (the EXPLETIVE) do-PRES NP<sup>x</sup> think Pro<sup>i</sup> be-PRES*. Nattinger (1980) gives examples such as *Down with . . .*. The term we use for such expressions is 'template form'.<sup>18</sup>

The *base pattern* of a lexicalised expression is the morpho-syntactic structure of its base form, expressed in terms of syntactic functions, categories and features.<sup>19</sup> For instance, *sitting duck* has the following base pattern:

(31) DETERMINER + PREMODIFIER (non-gradable adjective) + HEAD (singular count noun)

Because the base pattern is just the morpho-syntactic structure of an expression, several different base *forms* can have the same base *pattern*, e.g.:

(32) *a sitting duck*

(33) *a standing joke* ('a joke or humorous subject that is so well-known to a group of people that they laugh whenever it is mentioned')

(34) *a shooting star* ('a small piece of material from space which burns brightly as it passes through the Earth's air')

We use the term 'variant form' to refer to a form of a lexicalised expression which is not the base form. To be able to refer to a single morphological or syntactic alteration on the base form we use the term 'variation'. Thus the following variant form (in italics) contains only one variation (under-scored).

(35) *If you asked the man in the modern street* for his opinion of homosexuality, he would probably reply, 'I've nothing against queers . . .' (Birmingham Collection of English Texts; see Renouf 1984)

<sup>18</sup> We do not regard base forms with open initial determiner slots as template forms (e.g. DETERMINER *lost cause*).

<sup>19</sup> See Aarts & Aarts (1982) for a descriptive model of functions and categories.

And the following two variant forms (in italics) contain several variations (underscored):<sup>20</sup>

- (36) In frank terms, what has been revealed about the current scandals involving Kyowa and Sagawa are *only the tips of two of many icebergs floating in Japanese political waters*.
- (37) ... a deal to supply five aircraft to Aeroflot as *a hard-fought foot in the door of one of the aviation world's least exploited markets*.

Some expressions have more than one institutionalised form: one base form and one or more forms which are longer. These larger forms we call 'lexicalised variant forms'. In dictionaries the parts of these forms that are not required for the base forms are usually put between brackets. Two examples are:

- (38) *a bird in the hand (is worth two in the bush)*
- (39) *a hair of the dog (that bit you)*; ('an alcoholic drink taken in the morning because it is said to cure illness caused by drinking too much alcohol the night before')

We distinguish between four types of variation: permutation, term selection, addition and interruption. 'Permutation' implies that elements that are present in the base form are moved in the construction; their functions remain more or less the same.<sup>21</sup> Examples are:

- (40) *heir presumptive* vs. *presumptive heir*
- (41) *the sum total* vs. *the total sum*

In cases like these it is difficult to decide which form is the base form. In relation to the flexibility of these expressions it is more important which variation is possible (in this case permutation) than what the direction of this variation is (from pre- to postmodifier or vice versa).

'Term selection' implies that an element from a closed system is replaced with an element from the same system. Examples are:

- (42) *sitting duck* vs. *sitting ducks*
- (43) *the straw that breaks the camel's back* vs. *the straw that broke the camel's back*
- (44) *do something by fits and starts* vs. *do something in fits and starts*
- (45) *to be at a crossroads* vs. *to be at the crossroads*

'Addition' is the introduction of modifying or quantifying elements or of elements that do not interrupt the syntactic structure of the construction:<sup>22</sup>

<sup>20</sup> These examples are from *The Guardian* 1992.

<sup>21</sup> An example is modification. For instance, a premodifying adjective is used instead of a postmodifying adjective.

<sup>22</sup> See Nicolas (1992) for a semantic classification of additional modifiers in lexicalised expressions.



- (46) *a ruling passion* vs. *a ruling passion to discover new facts*
- (47) *a flash in the pan* vs. *a flash in the economic pan*
- (48) *the tip of the iceberg* vs. *the tip as well as the rest of the iceberg*

The fourth type of variation is 'interruption'. It is the insertion of a foreign element (that is not an immediate constituent) into the base form of a lexicalised expression:

- (49) (It was) *the talk of the town* vs. (It was) *the talk, yesterday, of the town*  
(interruption by adverbial)

We use the term 'flexibility' to refer to the extent to which constructions can take different grammatically possible forms. 'Free' constructions can take *all* forms that are grammatically possible. An example is:

- (50) *a book and a pen*

This construction can take forms like the following:

- (51) *a pen and a book* (permutation of NPs)
- (52) *books and pens* (term selections: plural nouns instead of singular ones)
- (53) *an expensive book and a cheap pen* (addition of premodifiers)
- (54) (They bought) *a book yesterday, and a pen too* (interruption by an adverbial)

It is evident that one would need several pages to discuss all grammatically possible forms. However, there is no reason to repeat here what various handbooks of grammar have described quite thoroughly already.

The most important characteristic of lexicalised expressions is that they cannot take all forms that are grammatically possible. To put it differently: when they are subjected to free grammatical variation, this variation will yield several pseudo-variants, i.e.: forms that can only be counterfeit forms.<sup>23</sup> Some lexicalised expressions have pseudo-variants only. Take for example the base form of the lexicalised expression *man and boy*. Grammatically possible pseudo-variants are:

- (55) # *men and boy* (term selection: plural form)
- (56) # *boy and man* (permutation: conjoins)
- (57) # *friendly man and naughty boy* (addition: two premodifiers)
- (58) # *the man and the boy* (addition: two central determiners)

On the other hand there are lexicalized expressions that have a limited number of variant forms, for instance: *a pain in the neck*. Some examples of variant forms are:

- (59) *a great pain in the neck*
- (60) *a pain in the neck of the trade unions*

<sup>23</sup> In Barkema (1994a) we used the terms 'unrelated variant' for pseudo-variant and 'related variant' for variant form.

Examples of pseudo-variants are:

(61) # . . . a pain, yesterday, in the neck

(62) # a pain in a neck

The various examples we have given in this section show that there are three types of construction: ones that have base and variant forms, ones that have base forms as well as variant forms and pseudo-variants, and ones with only base forms and pseudo-variants. We use the term *fully flexible*, *semi-flexible* and *inflexible constructions* respectively to refer to these three types. We only discuss the internal flexibility characteristics of constructions and not external ones such as the functions that constructions can take.

In some cases nominal heads which form part of lexicalised expressions can be premodified by adjectives such as *proverbial* and *old* (see e.g. Nicolas 1992). However, the addition of such elements has nothing to do with flexibility. It can be seen as a form of self-editing that has no effect on the established meaning at all, but only stresses the fact that the expression is lexicalised. An example is:

(63) an edifice of decisions as fragile as the proverbial card house  
(Birmingham Collection of English Texts).

### 3.4. Collocability

We use the term 'collocability' to refer to the degree to which it is possible to substitute a lexical item from an open class in a construction with alternatives from the same class: thus a noun is substituted by other nouns, a verb by other verbs, etc.<sup>24</sup> These alternatives can be synonyms, near-synonyms and antonyms. If the collocability of a construction is restricted, this is a matter of arbitrariness, i.e. the restriction is not linguistically motivated. In other words: a combination of lexical items in a construction may be in accordance with the morphological, syntactic and semantic rules of the language and at the same time be unacceptable as a form of a lexicalised expression. For instance, *the summit of the iceberg* is perfectly grammatical, but it is a counterfeit form of *the tip of the iceberg* ('the small part of a problem that is obvious, when the problem is much more serious and widespread than it seems to be', COBUILD, Sinclair 1987).

The greater the difference in size is between the set of theoretically possible alternatives and the subset of *acceptable* ones, the more the collocability of a construction is limited.

In 'free' constructions all lexical items from open classes can be substituted by any number of alternatives from the same classes. This number

<sup>24</sup> See also Carter (1987), chapter II. Firth (1951:123) already used the term 'collocability' more than forty years ago.

can be extremely large, depending on the 'ranges' of the lexical items – see Carter (1987, chapter III) or Cowie (1981:226). Take for example:

(64) *He observes the woman*

Both items from open classes – *observes* and *woman* – can be substituted by any number of alternatives from the classes of verbs and nouns. For example, *observe* can be replaced with synonyms such as *notice* or *look at*, while *woman* can be replaced with ones like *lady* or *missus*. The same goes for the template form  $X^{\text{time reference}} Y \text{ after } X^{\text{time reference}} Y$ , where time references like the following can be used: *day*, *week*, *minute* and *year*. We label such constructions as 'collocationally open'.<sup>25</sup>

On the other hand, the lexical items from open classes in the majority of lexicalised expressions cannot be replaced with alternatives from the same class. Take for instance:

(65) *by and large* ('on the whole, in general'); (not e.g.: # *by and big*, and # *near* and large)

(66) *red tape* ('silly detailed unnecessary official rules that delay action'); (not: # *green tape*, *red ribbon*, etc.)

The alternatives in examples 65 and 66 (preceded by a #) we call 'pseudo-alternants': with these alternative synonyms, near-synonyms or antonyms they can no longer be regarded as forms of lexicalised expressions but should be labelled as 'standard' constructions (or in the case of *by and big* as ungrammatical) instead. We call constructions in which none of the lexical items from open classes can be replaced 'collocationally closed'.

In some lexicalised expressions one or more open-class items can be replaced with (nearly) synonymous or antonymous items from the same class. Some examples are:

(67) *shanks's mare/pony*, ('walking')

(68) *generally/strictly speaking*

In example 67 *shanks's mare* and *shanks's pony* are alternants, while *shanks's horse* is a pseudo-alternant. In the same way *generally speaking* and *strictly speaking* in example 68 are alternants, while *elaborately speaking* is a pseudo-alternant.

If a lexical item from an open class in a lexicalised expression is replaced by an antonym, the resulting expression is generally an antonym of the original expression too, e.g. *a good sailor* ('one who enjoys travel by water without being sick') versus *a bad sailor*.

The number of alternatives in expressions like 67 and 68 is limited in an arbitrary way to a few or to only one. The alternative lexical items in these expressions are subject to the same selection restriction rules that apply to the ones they substitute for. We label such expressions as 'collocationally

<sup>25</sup> E.g. Carter (1987:63) uses the term 'unrestricted collocation'.



limited' and such replacements by (near-)synonyms or antonyms 'regular substitutions'.<sup>26</sup> There are exceptions: in a small number of expressions one or more lexical items from open classes can be replaced by ones from the same class, but these alternatives are no (near-) synonyms or antonyms. The semantic relations between the lexical items are arbitrary. For example:

(69) *monkey business/tricks*

We call such expressions collocationally limited as well, and call such replacements 'irregular substitutions'.

Different open-class items in lexicalised expressions can have different degrees of collocability. For the sake of completeness, we use the terms 'partly' and 'entirely' in combination with collocability in the same way as we use them in relation to idiomaticity. For instance, the template form *X<sub>time reference</sub> Y after X<sub>time reference</sub> Y* is collocationally 'entirely open', because both lexical items from open classes can be freely substituted, while *shanks's mare/pony* is both 'partly limited' (*mare* can be replaced with *pony*) and 'partly closed' (*shanks's* cannot be substituted).

In a small number of cases it is possible to replace a whole expression by a synonymous one in a number of steps. An example is: from *near miss* to *narrow shave*: *near miss* → *near thing* → *close thing* → *close shave* → *narrow shave*. The same goes for *old stager* and *good hand*: *old stager* → *old hand* → *good hand*. The insoluble question in relation to dominoes like this is: when are two forms one and the same lexicalised expressions, and when are they two or more expressions which are synonymous?

With the examples given in this section so far, we do not want to suggest that all *non*-lexicalised constructions are collocationally open. In section 3.1 we gave the following example of the restricted construction:

(70) *to work/operate/run a machine* (not e.g.: \**to manage a machine*, \**to conduct a machine*)

Another example of such a collocationally limited construction is:

(71) *to place a bet* (not e.g.: \**to put a bet*, \**to set a bet*, \**to station a bet*, \**to pose a bet*)

Carter (1987, chapter II) refers to Firth's (1957) examples 'strong tea' vs. 'powerful tea' (see also Greenbaum 1970:9). Such restricted constructions do not require special attention in dictionaries if their meanings are completely inferred from the basic and/or derived senses of the lexical items and the syntactic relations in these constructions and if their pragmatic functions can be determined on the basis of the context in which

<sup>26</sup> See Carter (1987:63) for the term 'semi-restricted collocation'. His class of 'familiar collocation' seems to be more or less similar to that of 'semi-restricted' ones.

they are used. However, as they are subject to collocational limitation, they are discussed in combinatory dictionaries such as BBI (Benson et al. 1986).

With a construction such as *crashing bore*, the inevitable question is: is it a lexicalised expression? According to LDOELC, *crashing* has a derived sense, 'very great'/'complete'; on the other hand, however, it only has this meaning in combination with *bore* and *fool*, and not in e.g. \**crashing liar*, \**a crashing idiot*, \**crashing nincompoop*. In our opinion *crashing bore* is a partly idiomatic lexicalised expression.

A particular type of substitution can be illustrated with the following examples:

- (72) But since *the proof of the beer is in the drinking*, we suggest that you try a can or two of authentic Draught Bass in the comfort of your home (advertisement in *The Independent*, 23/5/1992)
- (73) They therefore feel they are entitled to sing, throw plates and *trip the light fandango* during your act if they so choose. (Dillie Keane, *Punch*, 25/2/1992).
- (74) *A chip off the old shock* (heading in *The Independent*, 9/5/1992; reference to Robert Hughes' *The Shock of the New*)

Although the lexicalised expression *the proof of the pudding is in the eating*, *trip the light fantastic* and *a chip off the old block* are collocationally closed, the underscored parts in the examples above show that under special circumstances a particular type of replacement is even possible in closed expressions.<sup>27</sup> The rules for this type of replacement seem to be the following: the form is not an institutionalized alternant, and the basic sense(s) of the alternative lexical item(s) correspond(s) with the meaning of the context outside the expression. The substitution is either regular (*beer/pudding*; *eating/drinking*) or irregular. When it is irregular, the alternative is similar to the lexical item it replaces in another way, for example it rhymes (e.g. *block/shock*); it is also possible that the two lexical items have the same morpheme(s) (e.g. *fantastic/fandango*). It does not seem to be essential that the alternative lexical item belongs to the same class as the one it substitutes for, as one of the examples shows: in *trip the light fantastic/fandango* an adjective is replaced with a noun. We call this type of substitution 'punning' and a form of a lexicalised expression containing such a pun: 'a pun alternant'.

The effect punning has on the meaning of a lexicalised expression is similar to that of a variation, as in both cases the meaning is modified. However, as one lexical item is replaced by another, the extent to which pun alternants of lexicalised expressions are possible is not indicative of their flexibility. As the selection of alternatives cannot be regarded as an established one, but strongly depends on the context in which the expres-

<sup>27</sup> See also Fernando & Flavell (1981:38).

sion is used, the extent to which pun alternants are possible is not indicative of its collocability either.

### 3.5. Other types of characteristics

So far we have described the three main linguistic types of characteristics that all constructions have: compositionality, collocability and flexibility. In addition, subsets of lexicalised expressions have other characteristics.

Although most lexicalised expressions have base patterns similar to those of 'standard' constructions, a relatively small set have idiosyncratic syntactic structures and therefore do not have counterfeit forms. We call these 'idiosyntactic constructions'. Some examples are: *no can do*, *long time no see*, *make do*, *come to think of it*, *trip the light fantastic*, *by and large*, *gin and it* and *easy does it*.<sup>28</sup> Cruse (1986:38) uses the term 'asyntactic idiom' for such expressions. The disadvantage of this term is that it suggests that they do not have any syntactic characteristics. However, an expression like *by and large* can still be regarded as a form of coordination, be it an idiosyncratic one.

Some expressions are given special attention in dictionaries and handbooks of grammar because they have specific pragmatic characteristics, e.g.: *Ladies and Gentlemen*, *Good Morning*, *How do you do?* and *Happy Birthday!*<sup>29</sup> Fillmore et al. (1988) use the term 'idiom with pragmatic point', while Altenberg & Eeg-Olofsson (1990) label them as 'interactive expressions'. We call such constructions with institutionalised pragmatic functions 'formulaic expressions'. On the other hand, examples of pragmatically fixed expressions which are not (consciously) interactive, but expressive are: *God knows*, *What a performance*, *Good Lord* and *poor thing* (see e.g. CGEL, §11.53). We call these 'exclamatory expressions'.

Although most lexicalised expressions are monosemous, some are polysemous, for example *blind alley*: (1) 'a little narrow street with no way out at the other end'; (2) 'if a method of working or thinking is leading you up a *blind alley* or down a *blind alley*, it is turning out to be useless or not leading to good results' (COBUILD).

Finally, a number of lexicalised expressions have specific semantic characteristics; an example is *lift a finger*, which can only be used in negative contexts. For a description of such 'negative polarity items': see Tovenia (1992). Louw (1993) uses the term 'semantic prosody' for the semantic context in which a word or construction is generally used.

<sup>28</sup> See Fillmore et al. (1988) for the term 'extra-grammatical idiom'. Carter (1987, chapter III) refers to irregular syntactic structures such as *to go one better*, *to hold true* and *to go it alone*.

<sup>29</sup> See Quirk et al. (1985), §§11.38, 11.54 and 19.48.



#### 4. Examples of multi-dimensional classifications

As we said in section 1, the differences between the multi-dimensional model introduced in this article and the traditional one-dimensional classifications are explicitness and precision, uniformity in description, and the possibility to describe both types and tokens. To further illustrate the difference between the two classificatory methods, we discuss five expressions (i.e. types) and one example from a corpus (i.e. a token). In addition, we show how the expressions are treated in the following dictionaries:

- COBUILD: see Sinclair (1987);
- LDOELC: see Summers (1992);
- ODOCIE: see Cowie et al. (1983);
- OEED: see Hawkins and Allen (1991).

(75) *the bottom/first/highest rung of the ladder* ('the lowest/etc. level in the organization');

Multi-dimensional classification:

lexicalised expression:

- entirely idiomatic;
- semi-flexible (e.g. some variant forms are possible, e.g. . . . *political ladder*);
- collocationally partly open (because of *bottom/first/highest/etc.*), partly closed (because of *rung* and *ladder*).

One-dimensional classifications:

- non-compositional metaphor (Wood 1986);
- idiom (v/d Linden 1992);
- figurative idiom (Cowie 1983).

Wood's definition of the term 'idiom' does not allow its application to this expression, because it is not collocationally closed (see section 2.1). The fact that the expression is lexicalised is implied by the various traditional terms. None of these terms show whether the expression is semi-flexible or inflexible.

The dictionaries provide the following information: in COBUILD the expression is not listed as such; *rung* has a derived sense and the expression is used in an example. This gives us the wrong impression that the expression is a 'standard' construction in which *rung* has a derived sense and *ladder* an extended sense. In LDOELC the expression is a subentry of *rung*, which makes sense. In both OEED and ODOCIE the expression is absent.

(76) *If I may say so*:

Multi-dimensional classification:

- lexicalised expression:
  - fully compositional;
  - inflexible;
  - collocationally closed;
- formulaic

Fully compositional lexicalised expressions as example 76 are generally ignored in the literature on idioms, but they are discussed in handbooks of grammar because they are formulaic. In CGEL (§8.124) the expression is labelled as 'style disjunct'. The fact that it is lexicalised, inflexible and collocationally closed is implied by the handbook, not described. Although example 76 is a lexicalised expression, none of the dictionaries pays any attention to it.

(77a) *blind alley* (1) ('a little narrow street with no way out at the other end'):

NB: this expression is polysemous: its second sense is discussed in example 77b.

Multi-dimensional classification:

- lexicalised expression:
  - partly idiomatic: because of *blind*;
  - semi-flexible (e.g. plural is possible);
  - collocationally closed;
- polysemous.

One-dimensional classifications:

- semi-idiom (Cowie et al. 1983);
- collocation (Wood 1986);
- idiomatic expression (v/d Linden 1992).

Both Cowie et al. (1983) and van der Linden (1992) indicate with their terminology that *blind alley* is a partly idiomatic lexicalised expression, while Wood (1986) shows that it is collocationally closed. The fact that it is semi-flexible into the bargain, however, is not expressed in their terminology.

In COBUILD the sense of example 77a is absent; only its second sense is described (see example 77b). In LDOELC and ODOCE the expression is a main entry. In OED it is a subentry of *blind*.

(77b) *blind alley* (2) ('if a method of working or thinking is leading you up a blind alley or down a blind alley, it is turning out to be useless or not leading to good results'; COBUILD):

Multi-dimensional classification:

lexicalised expression:

- entirely idiomatic;
  - semi-flexible (plural is possible);
  - collocationally closed;
- polysemous.

One-dimensional classifications:

- figurative idiom (Cowie et al. 1983);
- idiom (Wood 1986);
- idiom (v/d Linden 1992).

All terms indicate that the expression is idiomatic. In addition, Wood's (1986) label implies that the expression is collocationally closed. None of the labels indicate that it is semi-flexible.

In COBUILD the expression is listed as a main entry; the dictionary indicates that it is semi-flexible and that the noun is countable. LDOELC only indicates that *blind alley* (1) can be used figuratively, thereby implying *blind alley* (2) is a 'standard' construction. OED lists the expression as a subentry of *blind*, ODOCIE as a main entry.

- (78) *bed and breakfast* ('(a private house or small hotel that provides) a place to sleep for the night and breakfast the next morning for a fixed price');

Multi-dimensional classification:

lexicalised expression:

- pseudo-compositional;
  - inflexible;
  - collocationally partly limited (because of *breakfast/board*), partly closed (because of *bed*).
- polysemous.

One-dimensional classifications:

- binomial (Quirk et al. 1985);
- collocation (Wood 1986);
- collocation (Cowie et al. 1983).

None of these classifications shows that the expression is lexicalised, pseudo-compositional and inflexible. The fact that it is collocationally limited is expressed by the classifications of Wood (1986) and Cowie et al. (1983), and implied by Quirk et al.'s (1985).

COBUILD gives the wrong impression that the expression (main entry) is monosemous. It is correct in indicating that it is semi-flexible and that the nouns are uncountable. LDOELC regards the expression as polysemous and provides two meanings. In ODOCIE the expression is absent, which makes sense in view of the fact that the dictionary focuses on



'idiomatic' English. OED, finally, has entered the expression as a subentry of *bed*. It is polysemous with three meanings; the first two meanings have to do with a night's lodging or the establishment where one can lodge, while the third is: 'an operation in which a shareholder sells a holding one evening while agreeing to buy it back again next morning, realizing either a gain or a loss in order to suit a tax requirement'.

(79) *What/ Why/ Who the X<sup>imum</sup>* ('used with various words, such as *hell*, *devil*, *blazes*, etc. when asking angry or surprised questions');

Multi-dimensional classification:

lexicalised expression:

- partly idiomatic (because of *the X<sup>imum</sup>*);
- inflexible;
- collocationally open (because of *X<sup>imum</sup>*). NB: the other items are from closed classes;

idiosyntactic;

template form;

exclamatory.

One-dimensional classifications:

colligation (Wood 1986);

exclamatory phrase (see CGEL, §11.53); emphasizer (CGEL, §8.100); also given as an example of a postmodifier of *wh*-words (CGEL, §7.64);

expletive (CGEL, §18.59n).

Wood's (1986) classification indicates that the expression is a template form, but does not express that it is partly idiomatic, inflexible, collocationally open and idiosyncratic. Because it is a collocationally open template form, it is typically an expression that should receive extensive attention in handbooks of grammar. Unfortunately, CGEL only pays attention to the exclamatory nature of the expression, while it ignores all other (idiosyncratic) characteristics.

For dictionaries the expression is problematic because it is collocationally open, so that none of the lexical items can be chosen as 'anchor'. In COBUILD the expression is listed as a subentry of *hell* (*why the hell*, *who the hell*, *what the hell*), of *blazes* (*the blazes*) and of *dickens* (*the dickens*). The dictionary indicates that the expression is inflexible. In LDOELC it is listed as a subentry of *what*, but not of *who* or *why*, etc. OED lists it as a main entry (*what the dickens*) and as subentries of *blaze* (*what the blazes*), *hell* (*what the hell*) and *devil* (*the devil*). In ODOCE the expression is a main entry (*what the hell/devil/blazes/heck/dickens*), while it is indicated that *what* can be replaced by *why*, etc.

So far, we have given examples of the characteristics of expressions, i.e. of *types*. On the basis of one example we show that our descriptive model can also be used for discussions of the characteristics of tokens:

- (80) And actually researching the current situation with *floating exchange rates* with all the differences that have cropped up over the past few years, there's very little written about it (Birmingham Collection of English Texts, spoken part).

Type: *rate of exchange*

lexicalised expression:

- pseudo-compositional;
- collocationally closed;
- semi-flexible (e.g. *rate* can be pluralised).

Token: *floating exchange rates*

base form: *rate of exchange*;

variant form: *floating exchange rates*:

- three variations: — one permutation: premodifier *exchange* vs. postmodifier of *exchange*;
- one addition; premodifier *floating*;
  - one term selection: pluralised head of the noun phrase (*rates*).

## 5. Conclusion

In the introduction to this article we discussed the disadvantages of the traditional terminology, which only highlights one striking feature of a lexicalised expression, and leaves various others implicit. We concluded that for detailed examination and discussion of the various linguistic dimensions of lexicalised expressions a well-defined model is required that distinguishes between various descriptive dimensions and at the same time pays heed to the scalar nature of the different types of characteristics.

In the remainder of the article we described a multi-dimensional model with which lexicalised expressions can be given as many labels as are required for exhaustive descriptions of their characteristics. In section 4 we illustrated the descriptive power of the model with multi-dimensional classifications of five types and one token.

The model can be used for classifications and descriptions in the literature, in handbooks of grammar and in dictionaries. Finally, it can be used for the annotation of the characteristics of lexicalised expressions in text corpora and for corpus-linguistic research into the relations between these characteristics. It was for this very purpose that the model was designed (see Barkema 1993, 1994a, 1994b, 1996).

## References

- AARTS, F. & AARTS, J., 1982. *English syntactic structures. Functions and categories in sentence analysis*. Oxford: Pergamon.
- AARTS, J. & CALBERT, J. 1979. *Metaphor and non-metaphor. The semantics of adjective-noun combinations*. Tübingen: Max Niemeyer.

- AARTS, J. & MEIJS, W. (eds.). 1984. *Corpus linguistics I. Recent developments in the use of computer corpora in English language research*. Amsterdam, Rodopi.
- AARTS, J. & MEIJS, W. (eds.) 1986. *Corpus linguistics II. New studies in the analysis and exploitation of computer corpora*. Amsterdam: Rodopi.
- AARTS, J. 1991. Intuition-based and observation-based grammars. *English corpus linguistics*, eds. K. Aijmer & B. Altenberg, 44–63. London: Longman.
- AARTS, J. & MEIJS, W. (eds.). 1990. *Theory and practice in corpus linguistics*. Amsterdam: Rodopi.
- AARTS J., DE HAAN, P. & OOSTDIJK, N. (eds.), 1993. *English language corpora: design, analysis and exploitation*. Amsterdam: Rodopi.
- AIJMER, K. & ALTENBERG, B. (eds.). 1991. *English corpus linguistics*. London: Longman.
- AISENSTADT, E. 1979. Collocability restrictions in dictionaries. *Dictionaries and their users*, ed. Hartmann, R. 71–4. Exeter.
- AISENSTADT, E. 1981. Restricted collocations in lexicology and lexicography. *ITL. Review of applied linguistics*, 53, 53–61.
- ALEXANDER, R. 1987. Problems in understanding and teaching idiomaticity in English. *Anglistik & Englischunterricht*, 32. *Wortschatzarbeit*. Heidelberg: Carl Winter.
- ALTENBERG, B. & EEG-OLOFSSON, M. 1990. Phraseology in spoken English: presentation of a project. *English corpus linguistics*, ed. J. Aarts & W. Meijs. Amsterdam, Rodopi.
- ALTENBERG, B. 1993. Recurrent verb-complement constructions in the London-Lund corpus. *English language corpora: design, analysis, and exploitation*, ed. J. Aarts & W. Meijs.
- BÄCKLUND, U. 1976. Frozen adjective-noun collocations in English. *Cahiers de lexicologie*, 28, 74–88.
- BARHEMA, H. 1993. Idiomaticity in English NPs. *English language corpora: design, analysis, and exploitation* ed. Aarts et al. 257–78.
- BARHEMA, H. 1994a. Determining the syntactic flexibility of idioms. *Creating and using English language corpora* ed. U. Fries, G. Tottie & P. Schneider, 39–52.
- BARHEMA, H. 1994b. The idiomatic, syntactic and collocational characteristics of received NPs: some basic statistics. *Hermes journal of linguistics*, 13:19–40. Århus: Århus School of Business.
- BARHEMA, H. 1996. The effect of inherent and contextual factors on the grammatical flexibility of idioms. *Synchronic corpus linguistics. Papers from the sixteenth international conference on English language research on computerized corpora*, Toronto, 1995 ed. I. Lancashire, C. Meyer & C. Percy, 217–31. Amsterdam: Rodopi.
- BENSON, M., BENSON, E. & ILSON, R. 1986. *The BBI combinatory dictionary of English*. Amsterdam: John Benjamins.
- BURGESS, A. 1969. Language made plain, *The English Language* ed. W. Boulton & D. Crystal, Vol. II. London: Cambridge U. Press.
- CARTER, R. 1987. Vocabulary: applied linguistic perspectives. *Aspects of English* ed. J. Sinclair. London: Allen & Unwin.
- CHAFE, W. 1970. *Meaning and the Structure of Language*. Chicago: U. of Chicago Press.
- COWIE, A. 1981. The treatment of collocations and idioms in learner's dictionaries. *Applied Linguistics*, 2, 223–35.
- COWIE, A., MACKIN, R. & McCAIG, I. 1983. *Oxford dictionary of current idiomatic English*. Oxford: Oxford U. Press.
- CRUSE, D. 1986. *Lexical semantics*. Cambridge: Cambridge U. Press.
- EVERAERT, M. & VAN DER LINDEN, E. 1989. *Proceedings of the first Tilburg workshop on idioms*. Tilburg: ITK Proceedings.



- EVERAERT, M. 1989. Verbal idioms and the lexicon. *Proceedings of the first Tilburg workshop on idioms* ed. M. Everaert & E. van der Linden, 1989, 57–78.
- EVERAERT, M. 1993. Vaste verbanden (in woordenboeken). *Spektator* 22, 1.
- EVERAERT, M., E. VAN DER LINDEN, A. SCHENK, & R. SCHREUDER (eds.) 1992. *Proceedings of idioms*, Tilburg: ITK-publication.
- EVERAERT, M., E. VAN DER LINDEN, A. SCHENK & R. SCHREUDER (eds.) 1995. *Idioms: structural and psychological perspectives*. Hillsdale (New Jersey): Lawrence Erlbaum Associates.
- FERNANDO, C. & FLAVELL, R. 1981. On idiom: critical views and perspectives, *Exeter linguistic studies*, 5, ed. R. Hartmann. Exeter: U. of Exeter.
- FILLMORE, C., KAY, P. & O'CONNER, M. 1988. Regularity and idiomaticity in grammatical constructions: the case of *let alone*. *Language* 64, 501–38.
- FIRTH, R. 1951. Modes of meaning. *Essays and studies*, 4. The English Association.
- FRASER, B. 1970. Idioms within a transformational grammar. *Foundations of Language* 6, 22–42.
- FRASER, B. 1976. *The verb-particle combination in English*. New York: Academic Press.
- FRIES, U., G. TOTTIE & P. SCHNEIDER (eds.) 1994. *Creating and using English language corpora. Papers from the fourteenth international conference on English language research on computerized corpora*. Zürich 1993. Series: Language and Computers, 13. Amsterdam: Rodopi.
- GAZDAR, G., KLEIN, E., PULLUM, G. & SAG, I. 1985. *Generalized phrase structure grammar*. Cambridge, Mass.: Harvard U. Press.
- GEERAERTS, D. 1992. Specialisation and reinterpretation in idioms. *Idioms: structural and psychological perspectives*, ed. Everaert et al.
- GREENBAUM, S. 1970. *Verb-Intensifier collocations in English*. The Hague: Mouton.
- HAWKINS, J. & ALLEN, R. 1991. *The Oxford encyclopedic English dictionary*. Oxford: Clarendon Press.
- HEALEY, A. 1968. English idioms. *Kivung*, 1, 71–108.
- HOCKETT, C. 1958. *A course in modern linguistics*. New York: Macmillan.
- JESPERSEN, O. 1924. *The philosophy of grammar*. London: Allen & Unwin.
- KATZ, J. & POSTAL, P. 1963. Semantic interpretation of idioms and sentences containing them. *MIT Quarterly Progress Report*, 70: 275–82.
- LOUW, B. 1993. Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. *Text and technology* ed. M. Baker, G. Francis & E. Tognini-Bonelli. Amsterdam: John Benjamins.
- MAKKAI, A. 1972. *Idiom structure in English*. The Hague: Mouton.
- MEL'ČUK, I. 1995. Phrasemes in language and phraseology in linguistics, *Idioms: structural and psychological perspectives* ed. M. Everaert et al. 167–233.
- MOORE, T. & CARLING, C. 1982. *Language understanding: towards a post-Chomskyan linguistics*. New York: St. Martin's Press.
- NATTINGER, J. 1980. A lexical phrase-grammar for ESL. *TESOL Quarterly*, 14, 3, 337–44.
- NEWMAYER, F. 1974. The regularity of idiom behavior. *Lingua* 34, 307–42.
- NICOLAS, T. 1992. Internal modification of English V-NP idioms. *Idioms: structural and psychological perspectives*. ed. Everaert et al. 85–96.
- NICOLAS, T. 1995. Semantics of idiom modification. In: Everaert et al. (eds.) 1995: 223–52.
- PAWLEY, A. & SYDER, F. 1983. Two puzzles for linguistic theory: nativelike selection and nativelike fluency. *Language and communication*, ed. J. Richards & R. Schmidt, 191–227. London: Longman.
- QUIRK, R., GREENBAUM, S., LEECH, G. & SVARTVIK, J. 1985. *A comprehensive grammar of the English language*. London: Longman.

- RENOUF, A. 1984. Corpus development at Birmingham U. *Corpus linguistics II*, ed. J. Aarts & W. Meijs. 1984: 3–39.
- SAUSSURE, F. DE. 1916. *Cours de linguistique générale*. Paris: Payot. Translation: *Course in general linguistics*. London: Fontana. 1974.
- SIMPSON, J. & E. WEINER. 1989. *The Oxford English dictionary*. Oxford: Clarendon Press.
- SINCLAIR, J. 1987. *Collins COBUILD English language dictionary*. Glasgow: Collins.
- SUMMERS, D. 1992. *Longman dictionary of English language and culture*. Harlow: Longman.
- SWEET, H. 1889. *The practical study of languages*. London: Oxford U. Press. Reprinted in 1964.
- TOVENA, L. 1992. Polarity sensitive idioms in translation. *Proceedings of idioms*, ed. Everaet et al. 127–41. Tilburg: ITK Proceedings.
- ULLMANN, S. 1973. *Semantics. An introduction to the science of meaning*. Oxford: Blackwell.
- VAN DER LINDEN, E. 1992. *Idioms, non-literal language and knowledge representation*. Tilburg: ITK research memo, 12.
- VAN DER VOORT, M. & W. VONK. 1995. You don't immediately die when you kick an empty bucket. *Idioms: structural and psychological perspectives*, ed. M. Everaert et al. 283–300.
- WEINREICH, U. 1969. Problems in the analysis of idioms. *Substance and structure of language*, ed. J. Puhvel. 23–81. Berkeley: U. of California Press.
- WILLIAMS, E. Unpublished manuscript. Remarks on lexical knowledge. Princeton U.
- WOOD, M. 1986. *A definition of idiom*. Indiana, Bloomington: Indiana U. Linguistics Club.

Henk Barkema  
TOSCA Research Group  
Dept. of Language and Speech  
Corpus Linguistics Section  
Faculty of Arts  
University of Nijmegen  
P.O. Box 9103  
6500 HD Nijmegen  
The Netherlands  
e-mail: barkema@let.kun.nl

## Appendix: alphabetical list of terms and their definitions

- *absent sense*: an *absent sense* expression is a lexicalised expression which contains at least one lexical item which does not have a (basic) sense in isolation (see also: *basic sense*).
- *alternant*: the different lexical realizations of the base form of one and the same collocationally limited or open lexicalised expression are its *alternants* (see also: *pseudo-alternant* and *pun alternant*).
- *base form*: a *base form* is the simplest form of an expression which contains the lexical items and exhibits the minimal syntactic structure required for the specific meaning and/or pragmatic function of the expression. A base form is always a constituent (see also: *template form*, *variant form* and *base pattern*).



- *base pattern*: a *base pattern* is the syntactic structure of a base form, expressed in terms of syntactic functions, categories and features (see also: *base form*).
- *basic sense*: the *basic sense* of a lexical item is the sense from which other senses can be systematically derived by means of extension. It is usually the first sense that comes to a speaker's mind when s/he comes across a lexical item in isolation (see also: *extended* and *derived sense*).
- *collocability*: the *collocability* of a construction is the degree to which it is possible to replace the open-class lexical items in the construction with (nearly) synonymous or antonymous items from the same class. The *degree* of collocability depends on the number of open class items in the construction that can be replaced as well as on the number of items that can take their places as alternatives (see also: *collocationally open*, *collocationally limited* and *collocationally closed*).
- *collocationally closed*: a *collocationally closed* construction is a construction in which *none* of the open-class items can be replaced with (nearly) synonymous or antonymous items from the same class; a construction can be collocationally entirely closed, but also partly (see also: *collocationally open*, *collocationally limited*, *entirely* and *partly*).
- *collocationally limited*: a *collocationally limited* construction is a construction in which at least one open-class lexical item can be replaced with at least one other (nearly) synonymous or antonymous item from this class; a construction can be collocationally entirely limited, but also partly (see also: *collocationally open*, *collocationally closed*, *entirely* and *partly*).
- *collocationally open*: a *collocationally open* construction is one in which *all* open-class items in the base form can be replaced with any number of (nearly) synonymous or antonymous items from the same class. Substitution is only subject to selection restrictions; a construction can be collocationally entirely open, but also partly (see also: *collocationally limited*, *collocationally closed*, *entirely* and *partly*).
- *compositionality*: the compositionality of a construction is the extent to which its meaning is the combinatorial result of the basic or derived senses of its lexical items and the syntactic structure of the constituent that contains these lexical items (see also: *fully compositional*, *pseudo-compositional*, *partly compositional* and *non-compositional*).
- *construction*: a construction is a phrase, clause or sentence with (a) at least two function slots, and in addition (b) at least two lexical items from open classes, or (c) at least one lexical item from an open class and a grammatical function word other than the definite or indefinite article (see also: *expression* and *standard construction*).
- *counterfeit form*: a *counterfeit form* is a standard construction with the same syntactic form and lexical items as the base form or a variant form of a lexicalised expression. However, it does not have the same idiosyncratic meaning and/or the same specific pragmatic function as this lexicalised expression (see also: *standard construction* and *lexicalised expression*).
- *derived sense*: a *derived sense* of a lexical item is every sense, other than the basic sense, that a word has in isolation (see also: *basic* and *extended sense*).
- *entirely*: an expression is *entirely idiomatic* or *collocationally entirely open/closed/limited* if the whole expression has this characteristic, and not only part of it (see also: *partly*).
- *exclamatory*: an *exclamatory* expression is an expression with an institutionalised expressive pragmatic function: it is used by the speaker to express strong feeling (see also: *formulaic*).
- *expression*: an *expression* is a familiar construction (see: *familiar construction* and *lexicalised expression*).



- *extended sense*: an *extended* sense of a lexical item is a sense which is the unique result of a violation of restriction rules that results in the activation of sense adaptation rules. The following factors play a role in the sense adaptation: (a) the basic sense of the lexical item, (b) the speaker's stereotyped beliefs about the referents of this basic sense and (c) the syntactic environment in which it occurs (see also: *basic* and *derived* sense).
- *familiar*: a *familiar construction* is one which has gone through a gradual process of institutionalisation. As a result it has become current within the language community with a specific meaning and/or pragmatic function (see also: *expression* and *lexicalised expression*).
- *flexibility*: the *flexibility* of a construction is the extent to which it can take different grammatically possible forms (see also: *fully flexible*, *semi-flexible* and *inflexible*).
- *formulaic*: a *formulaic* expression is an expression with an institutionalised interactive pragmatic function: it is addressee-oriented (see also: *exclamatory*).
- *free*: a construction is *free* if the generally applicable rules can be applied to it without any (idiosyncratic) restriction (see also: *idiosyncratic* and *restricted*).
- *fully compositional*: a *fully compositional* construction has a meaning that is entirely the combinatorial result of the senses of its lexical items and the syntactic structure of the constituent that contains these lexical items (see also: *pseudo-*, *partly-*, and *non-compositional*).
- *fully flexible*: a *fully flexible* construction is one which can take *all variations* that are grammatically possible (see also: *semi-flexible* and *inflexible*).
- *idiomatic expression/idiom*: an *idiomatic expression* or *idiom* is a lexicalised expression which is non-compositional (see also: *lexicalised expressions*, *entirely idiomatic* and *partly idiomatic*).
- *idiosyncratic*: restrictions on generally applicable rules or structures are *idiosyncratic* if they cannot be formulated in terms of (general) rules.
- *idiosyntactic*: an *idiosyntactic* construction is one with an idiosyncratic syntactic structure (see *idiosyncratic*).
- *inflexible*: an *inflexible* construction is one which can take *no variations* that are grammatically possible, i.e. it has pseudo-variants only (see also: *fully flexible* and *semi-flexible*).
- *lexicalised expression*: a *lexicalised expression* is a familiar construction with at least two idiosyncratic linguistic characteristics: it has limited flexibility and in addition it is collocationally limited and/or not fully compositional (see also: *expression* and *familiar construction*).
- *lexicalised variant form*: if a lexicalised expression has two or more institutionalised forms, we call the simplest form its *base form* and the longer form(s) its *lexicalised variant form(s)* (see also: *variant form*).
- *monosemous*: an expression is *monosemous* if it has only one institutionalised meaning (see also: *polysemous*).
- *non-compositional*: a *non-compositional* construction has a meaning that is not the combinatorial result of the senses of its lexical items and the syntactic structure of the constituent that contains these lexical items (see also: *fully compositional*, *pseudo-compositional*, *partly-compositional* and *(partly/entirely) idiomatic*).
- *partly*: an expression is *partly compositional*, *partly idiomatic* or *collocationally partly open/closed/limited*, if one part of the expression has this characteristic, and another part has another characteristic (see *entirely*).
- *polysemous*: an expression is *polysemous* if it has two or more institutionalised meanings (see also: *monosemous*).
- *pseudo-alternant*: if in the base form of a lexicalised expression one or more lexical items have been replaced by one or more (nearly) synonymous or antonym-

mous terms while this alternative does not have the same idiosyncratic meaning and/or pragmatic function as the lexicalised expression or its antonym, this construction is a *pseudo-alternant* (see also: *alternant* and *pun alternant*).

- *pseudo-compositional*: in a *pseudo-compositional* construction only part of its meaning is the combinatorial result of the senses of all of its lexical items and the syntactic structure of the constituent that contains these lexical items (see also: *fully compositional*, *partly compositional* and *non-compositional*).
- *pseudo-variant*: if a construction shows lexical resemblance to, and is grammatically a variant form of, a lexicalised expression, but does not have its idiosyncratic meaning and/or pragmatic function, it is a *pseudo-variant* (see also: *variation* and *variant form*).
- *pun alternant*: a *pun alternant* is a base or variant form of a lexicalised expression with one or more regular or irregular substitutions that are not established alternatives but puns (see also: *alternant* and *pseudo-alternant*).
- *restricted*: an expression is *restricted* if not all generally applicable grammar rules can successfully be applied to it. An expression can be restricted in its compositionality, its collocability and in its flexibility. These restrictions are generally idiosyncratic (see *idiosyncratic* and *free*).
- *semi-flexible*: a *semi-flexible* construction is one which can take a limited number of variations that are grammatically possible, i.e. it has both variant forms and pseudo-variants (see also: *fully flexible* and *inflexible*).
- *standard*: a *standard construction* is one which has been formed by means of the general rules of the grammar of a language (see also: *expression*, *free* and *counterfeit form*).
- *template form*: a *template form* is a base form with one or more open function slots (see also: *base form*).
- *token*: a *token* is a realisation of an expression in context (see also: *type*).
- *type*: a *type* is an expression, abstracted from its realisation in context (see also: *token*).
- *variant form*: a *variant form* is a form of a lexicalised expression with one or more variations that has basically the same idiosyncratic meaning and/or pragmatic function as the base form (see also: *variation* and *pseudo-variant*).
- *variation*: a *variation* is a single alteration in the base form of a lexicalised expression. It involves permutation, term selection, addition or interruption (see also: *variant form* and *pseudo-variant*).